Can Automated Responses to Cyber Incidents be supported by Artificial Intelligence?

Manuel Zambelli, PhD student in Advanced-Systems Engineering
Prof. Barbara Russo, Supervisor and Coordinator of the PhD program
Software Engineering and Autonomous Systems (SEAS)
Faculty of Engineering, Free University of Bolzano

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unibz
Südtiroler Informatik AG Informatica Alto Adige SPA

Outline

- Motivation(s)
- Log files
- Security Information and Event Management (SIEM)
- Security Orchestration, Automation, and Response (SOAR)
- Manual labelling for AI Integration
- Incident descriptions
- Generation of the incident description as a vector
- Conclusions

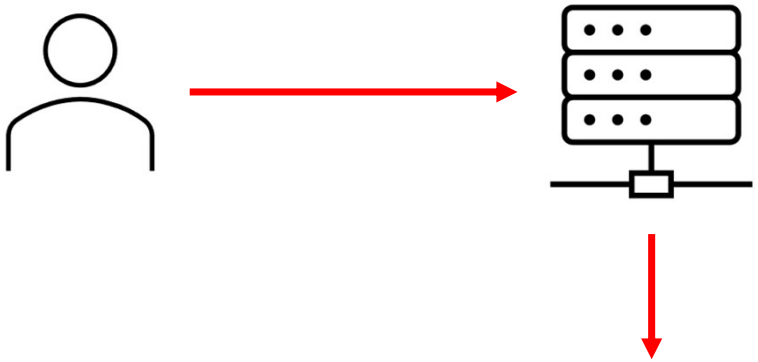
Motivation 1: False Positives

- Problem of **False Positives** (FP) in Machine Learning
- Cyber incidents which are **misclassified**
- **Increases the costs** in the company (more working hours)
- Examples:
  - User login after failed attempts
  - Unauthorized access which does not compromise the system or steal data

Motivation 2: Cyber Taxonomy

- **NIS2** (EU Directive 2022/2555, D.Lgs. n.138/2024)
- Starting from January 2026: communication of cyber incidents with a significant impact to the **Computer Security Incident Response Team** (CSIRT)
- The **Agenzia per la Cybersicurezza Nazionale** (ACN) defines a taxonomy (**TC-ACN**) for incident notifications:
  4 macrocategories, 22 predicates, 144 values

Log Generation in Network Devices

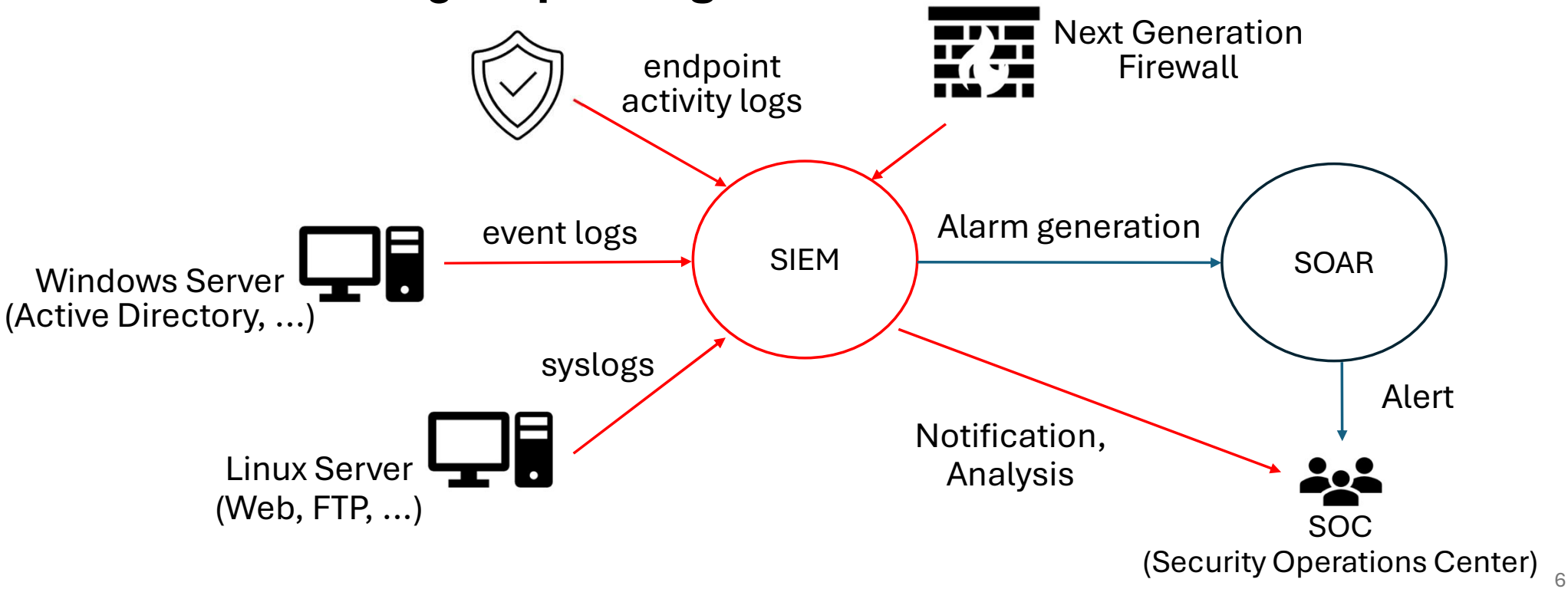


Überwachung erfolgreich	02.11.2025 11:34:59	4611	Security System Extension
Überwachung gescheitert	02.11.2025 11:34:59	4625	Logon
Überwachung erfolgreich	02.11.2025 11:34:59	5059	Other System Events
Überwachung erfolgreich	02.11.2025 11:34:59	5059	Other System Events
Überwachung erfolgreich	02.11.2025 11:34:59	4611	Security System Extension

SIEM

Security Information and Event Managament:

Collection of logs + **parsing**



SIEM solutions

Splunk Enterprise Security

Microsoft Sentinel

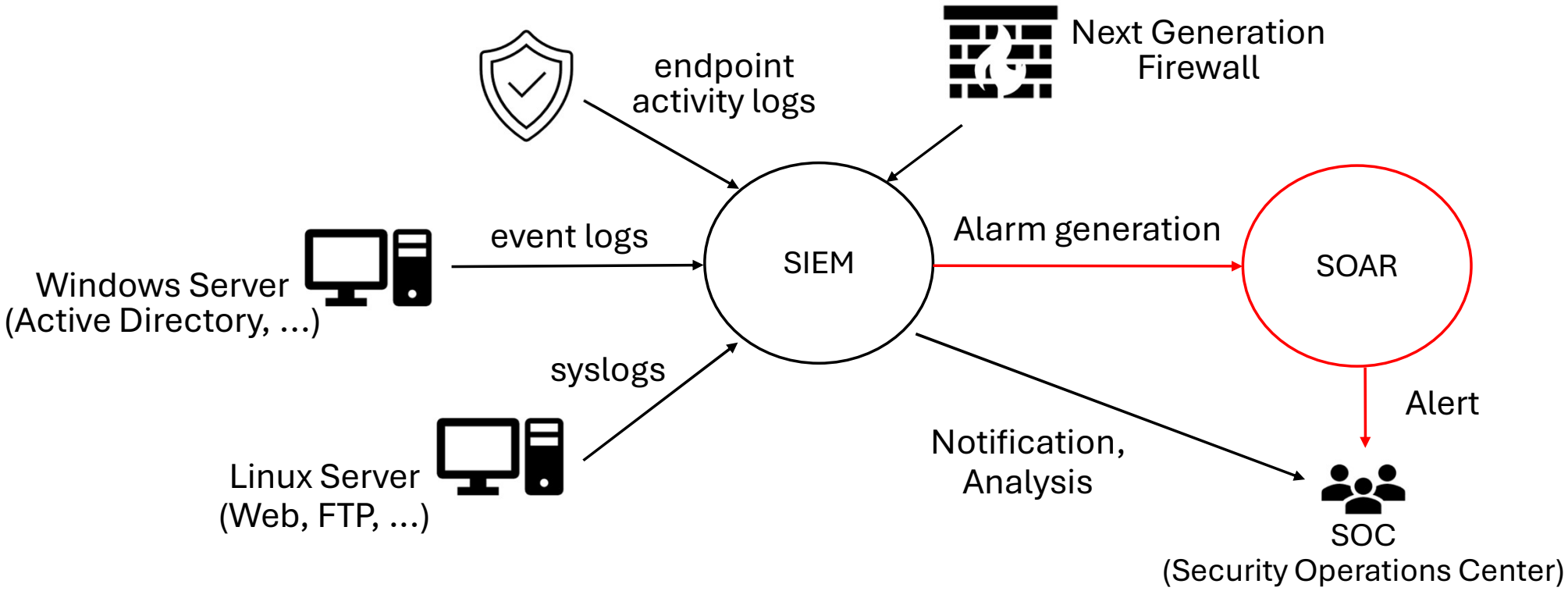
IBM QRadar

Elastic Security

Open source solutions (Elastic Stack, OSSIM, OSSEC, …)

SOAR

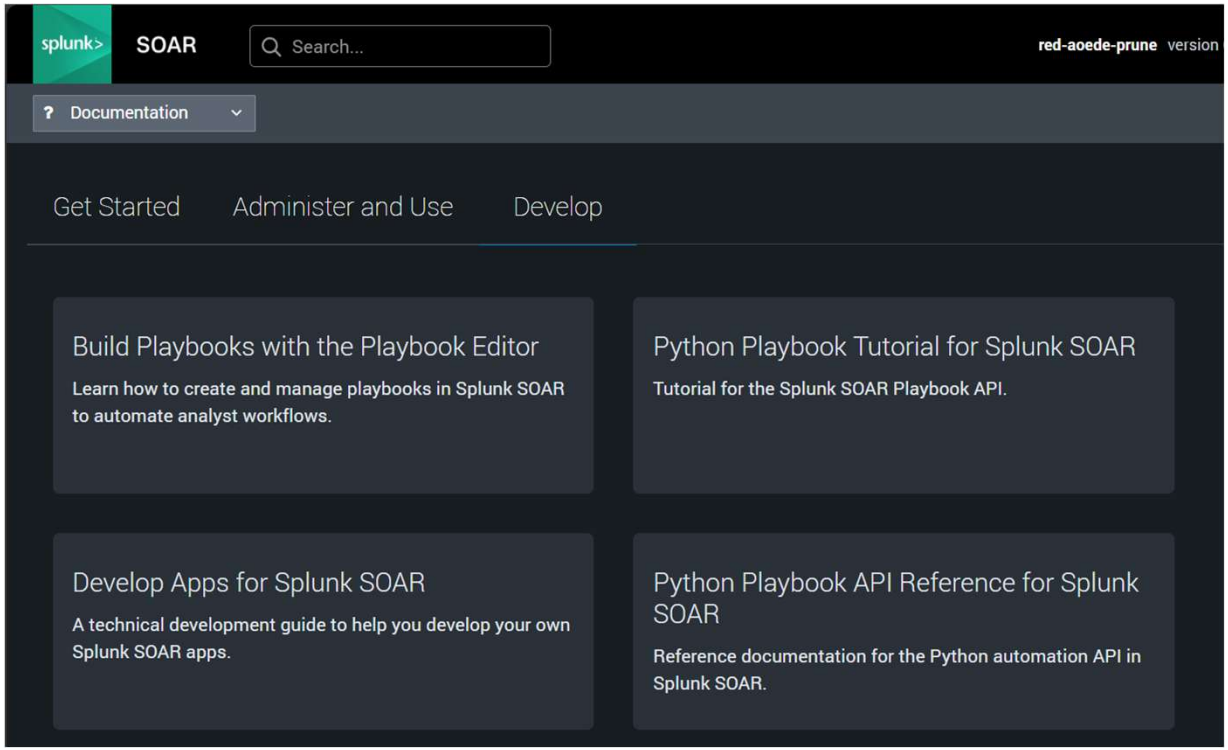
Security Orchestration, Automation, and Response:



Splunk SOAR

Setup:

- On-premises installation
- Free Community Edition
- Phantom Community Playbooks (Apache-2.0 license)

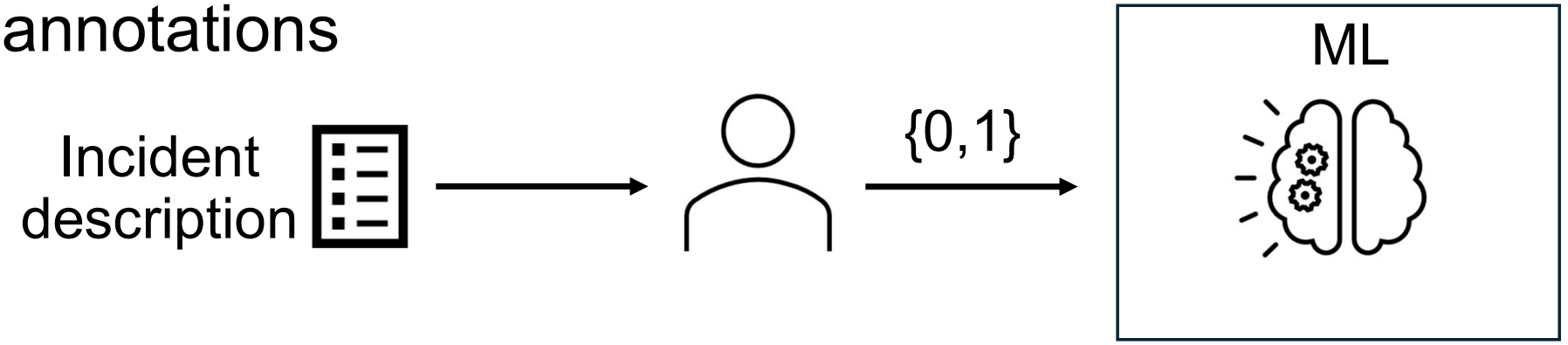


https://www.splunk.com/en_us/download/soar-free-trial.html

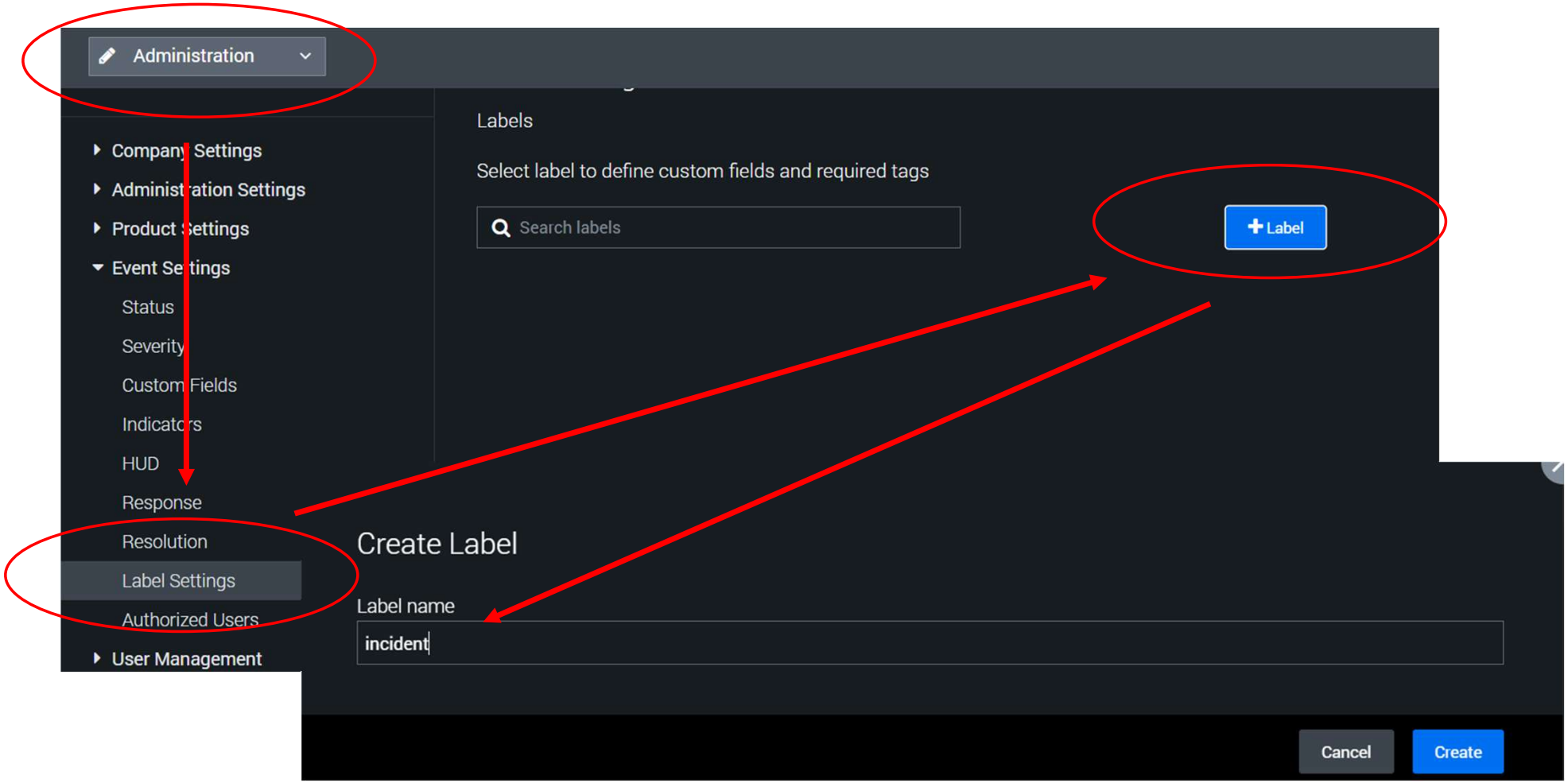
How to reduce the FP

Methodology

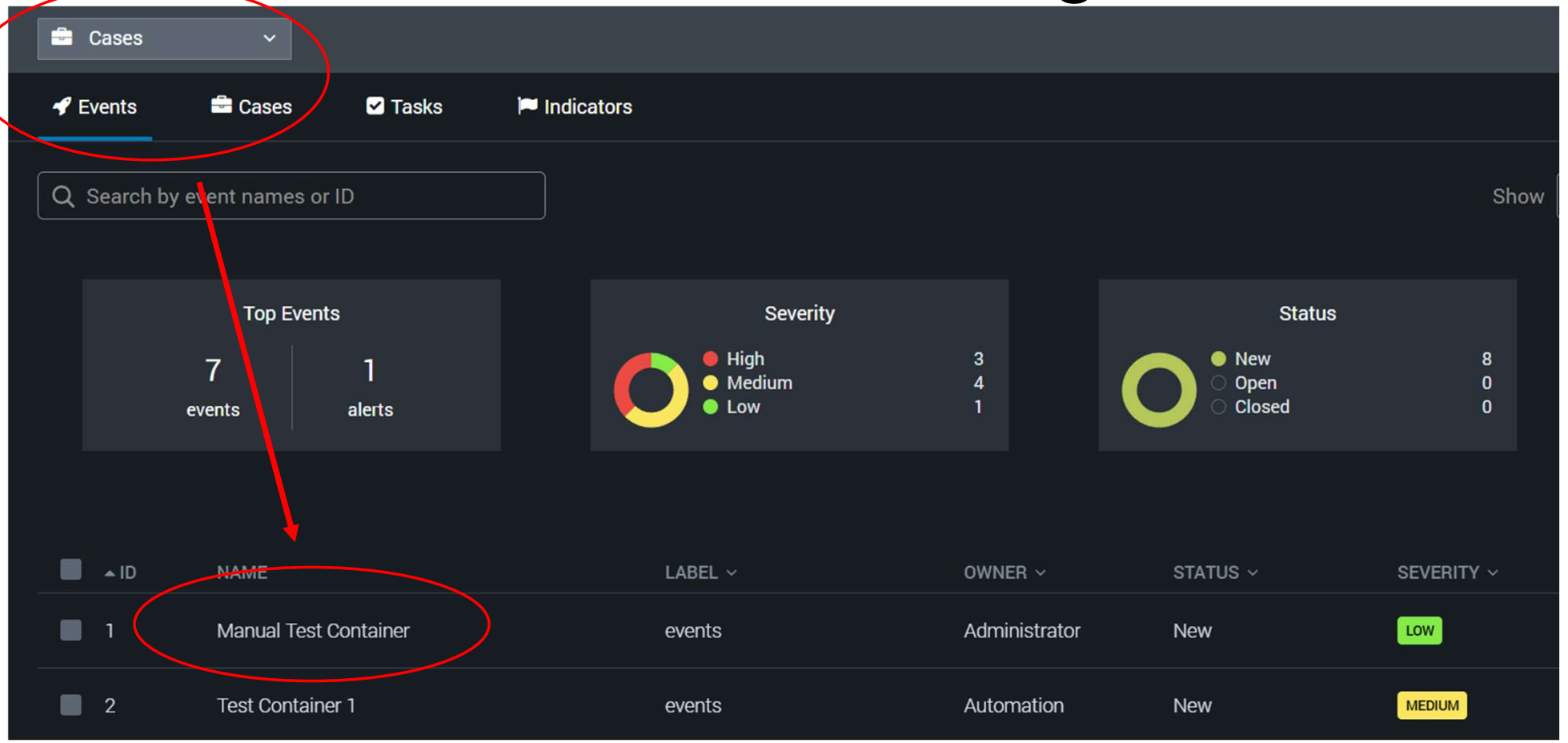
- Manual labelling → Ground Truth
- Training a Machine Learning (ML) model
- Continual Learning through human annotations



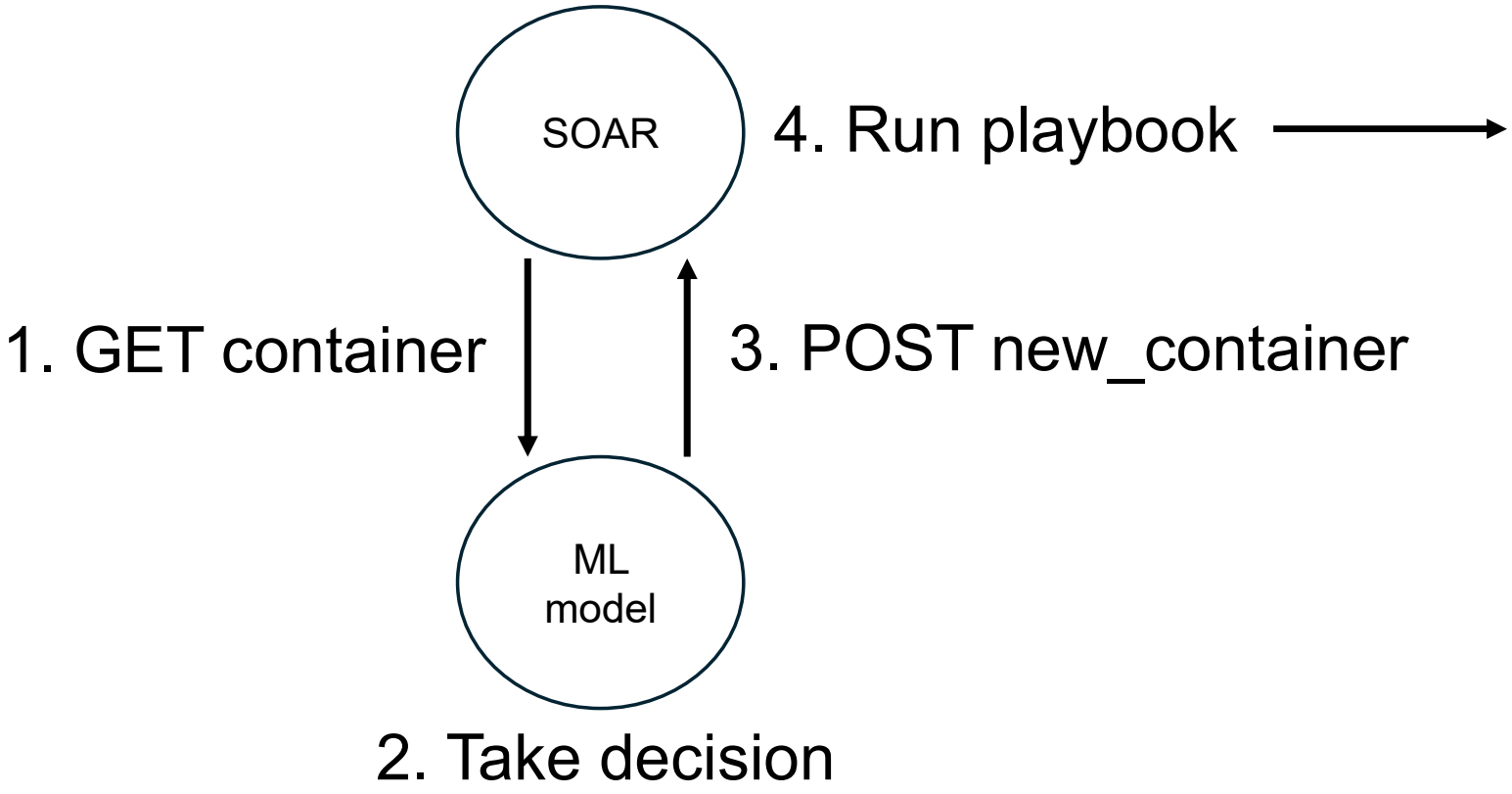
Create new label



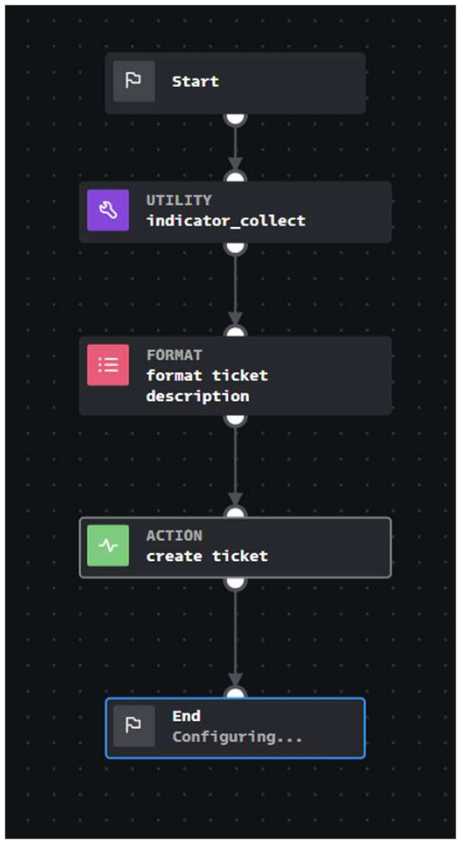
Manual Labelling



AI Integration



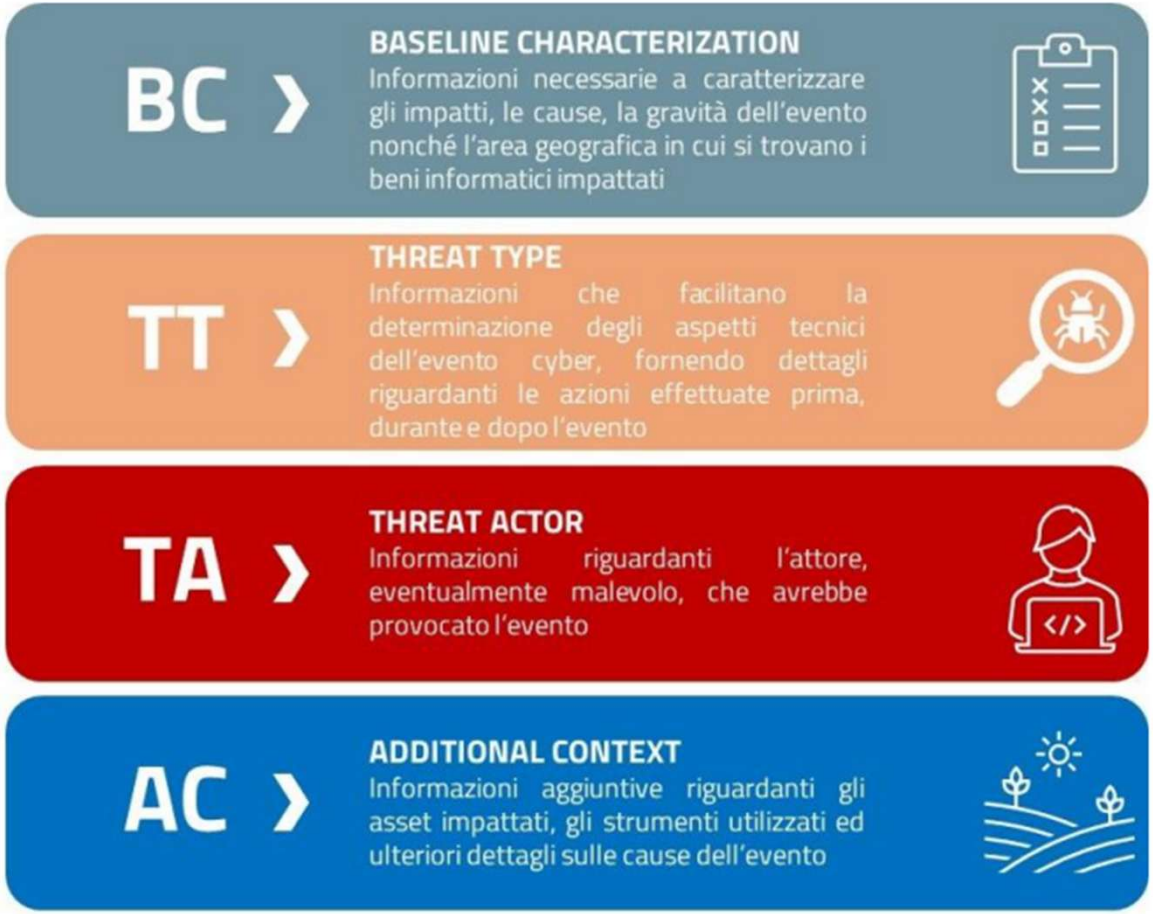
Visual Playbook Editor



Playbooks

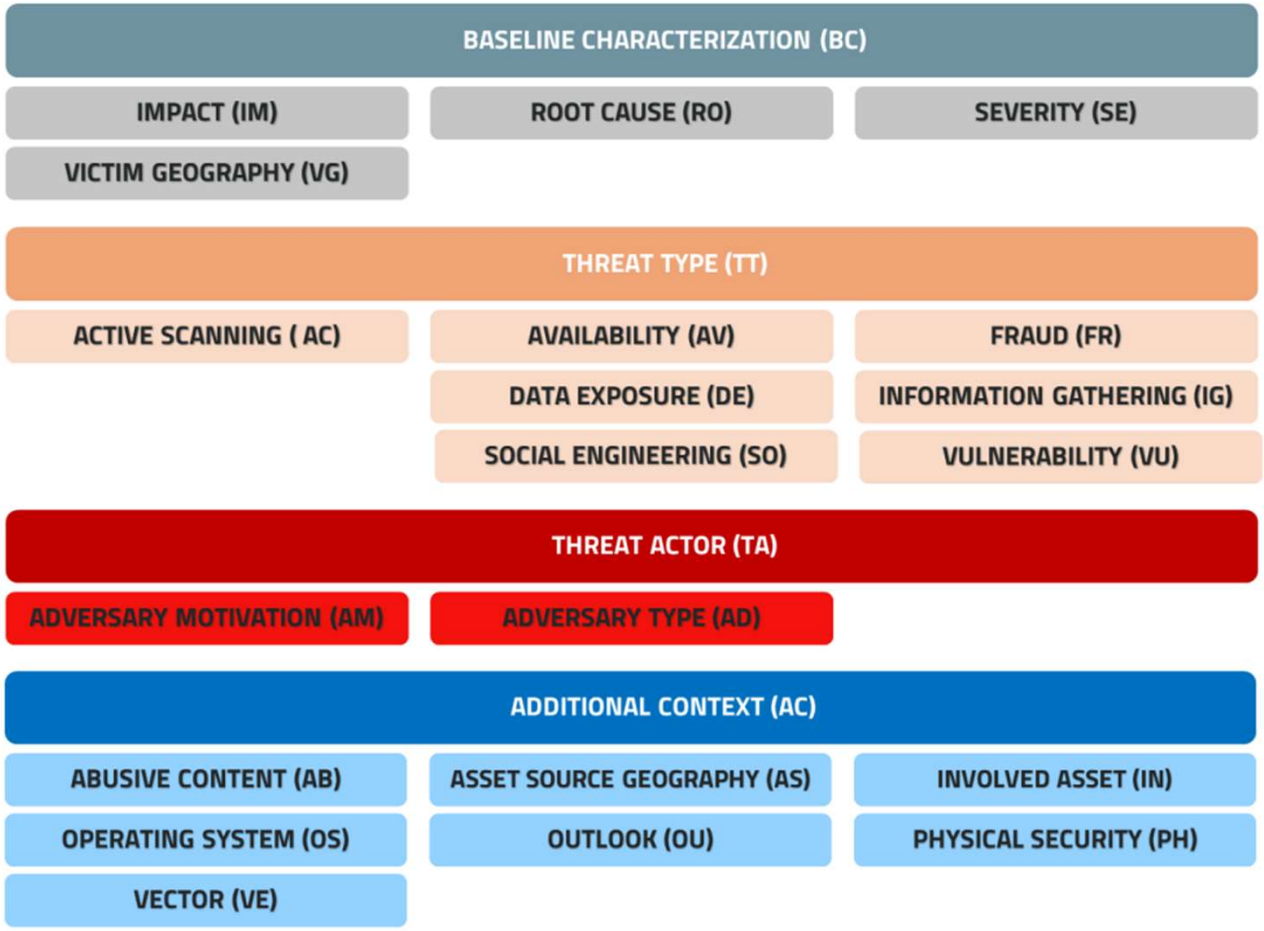
- Phantom Community Playbooks
- Phantom library for Python
- Splunk playbooks: https://research.splunk.com for the Visual Playbook Editor (VPE)
- Phantom Community Playbooks on github: https://github.com/phantomcyber/playbooks (Apache-2.0 license)

Incident reporting according to TC-ACN



https://www.acn.gov.it/portale/documents/d/guest/acn_tassonomia_cyber_clear

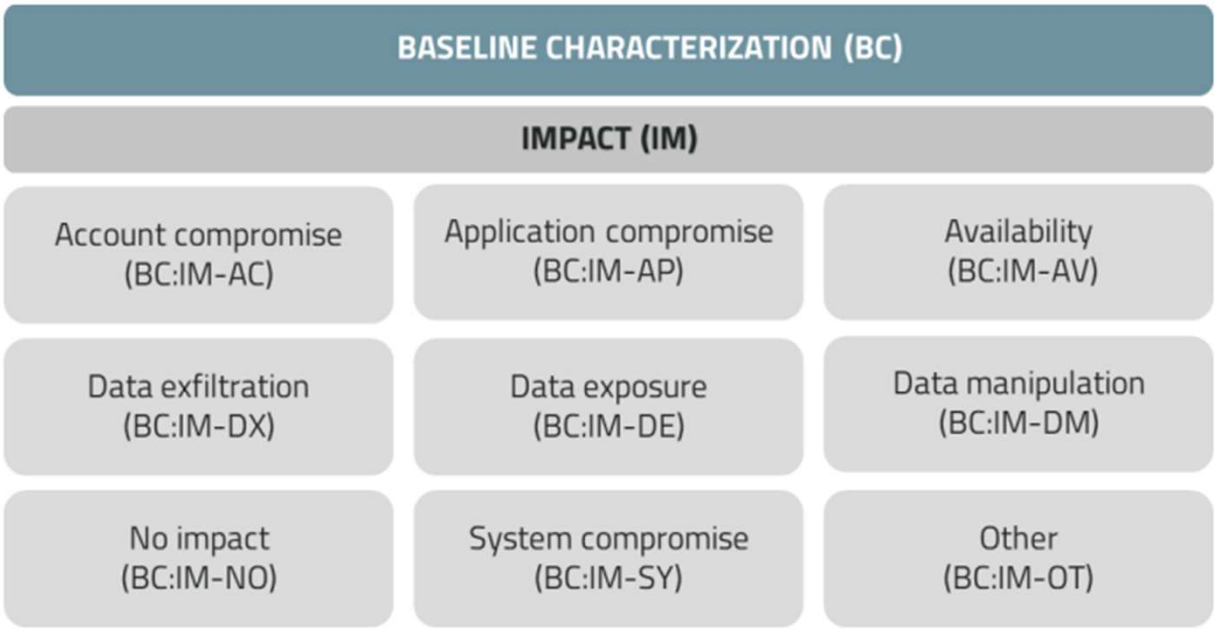
Macrocategories and Predicates



https://www.acn.gov.it/portale/documents/d/guest/acn_tassonomia_cyber_clear

Values

Values for the macrocategory **Baseline Characterization** and the predicate **Impact**:



https://www.acn.gov.it/portale/documents/d/guest/acn_tassonomia_cyber_clear

Example of TC-ACN vector

<BC:IM-DE BC:RO-MA BC:SE-LO BC:VG-EU
TT:DE-PD TT:MA-BA TT:VU-SE TA:AM-ES TA:AD-CR
AC:AB-OT AC:AS-IT AC:IN-OT AC:PH-UN>

→ Impact: Data Exposure

→ Root Cause: Malicious Actions

→ Severity: Low

→ Victim Geography: Europe

→ Data Exposure: Personal data

→ Malicious Code: Backdoor

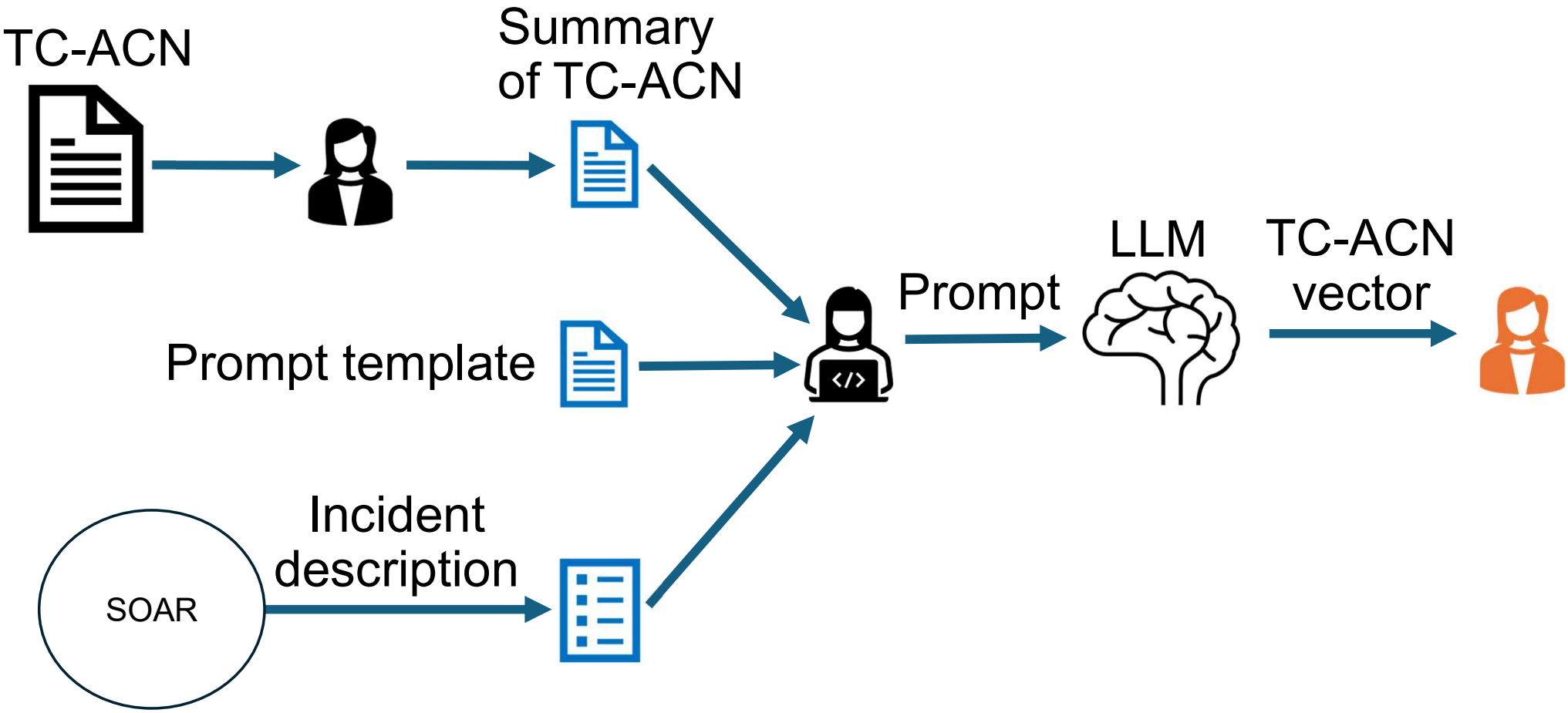
…

MITRE ATT&CK Framework

250 adversary tactics and techniques (https://attack.mitre.org)

https://attack.mitre.org

Methodology



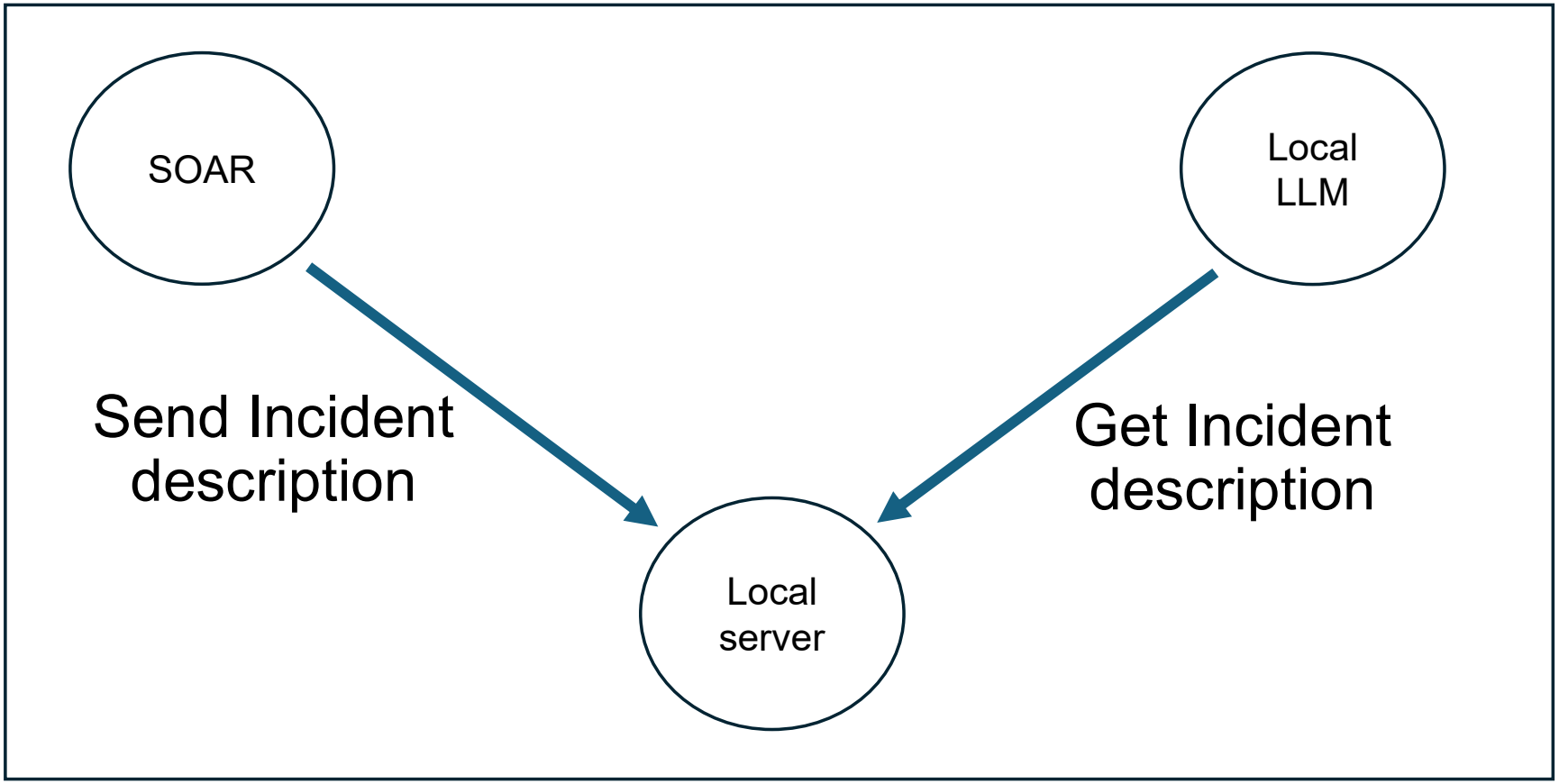
How to generate the prompt

Prompt:

You are part of a SOC team. Your role is the security specialist. Consider the incident description <incident description>. Generate the TC-ACN vector in the correct format according to the definition of the TC-ACN taxonomy <TC-ACN summary>.

Automation

On-premises solution



Preliminary Results

Setup: gpt-oss:20b model, thinking feature, 4k context length

Synthetic data for the incident description

Incident description: „Potential DoS Attack detected“ →

<BC:IM-AV BC:RO-MA BC:SE-ME BC:VG-GL TT:DS-OT AC:AB-OT AC:AS-OT AC:IN-OT AC:PH-OT TA:CY-OT>

Preliminary Results

Setup: gpt-oss:20b model, thinking feature, 4k context length

Synthetic data for the incident description

Incident description: „Web application exploited through SQL injection“ →

<BC:IM-AP BC:RO-MA BC:SE-ME BC:VG-GL TT:VU-OT TA:CY-OT AC:AV-UN AC:TS-WS AC:ET-UN AC:DM-UN AC:RA-UN>

Conclusions

- AI Augmentation to address the problem of FP
  (it needs a ground truth / human feedback)
- Phantom Community Playbooks to test automated responses
  (basic concepts can be applied to different software solutions)
- Brief introduction to TC-ACN
- Leveraging (local) LLMs
- Bias problem → Improving the quality of the TC-ACN summary
- Introducing AI models to replace humans
- Fine-Tuning LLMs with examples

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unibz
Südtiroler Informatik AG Informatica Alto Adige SPA

This work is co-funded by the Italian MIUR (with NRRP funds) and by the local company SIAG / Informatica Alto Adige S.p.A. with the scolarship provided under the project number

CUP I52B24000510005.

We would like to acknowledge the assistance and the support of Ing. Francesco Terracciano and Mr. Abdalrahman Hwoij from the company SIAG / Informatica Alto Adige S.p.A.

References and Links

- Agenzia per la cybersicurezza nazionale (2025), La direttiva NIS, https://www.acn.gov.it/portale/nis
- Agenzia per la cybersicurezza nazionale (2025), La Tassonomia Cyber dell'ACN, https://www.acn.gov.it/portale/documents/d/guest/acn_tassonomia_cyber_clear
- Ollama (2025), Ollama models, https://ollama.com/search
- Splunk playbooks (2025): https://research.splunk.com
- Phantom Community Playbooks on github (2025): https://github.com/phantomcyber/playbooks
- MITRE ATT&CK Framework (2025): https://attack.mitre.org

Note: All listed links were last accessed on 02/11/2025.